



## Report on the LOCKSS Beta Test



The mission of the Government Printing Office (GPO) is to provide permanent public access to official Federal Government publications in print and electronic formats through the Federal Depository Library Program (FDLP). In addition, the International Exchange Service (IES) operates according to 44USC1719 which states, "there shall be supplied to the Superintendent of Documents, Government publications for distribution to those foreign governments which agree, as indicated by the Library of Congress, to send to the United States similar publications of their governments for delivery to the Library of Congress."

Various distribution systems (e.g. tangible depository distribution, *GPO Access*, etc.) exist to facilitate dissemination via the FDLP and IES. A federated dissemination system exists for the management and distribution of Government journals in print format, but a similar federated system does not exist for journals in electronic format (e-journals). In the print environment, copies of a Government journal are distributed to Federal depository libraries around the country through the FDLP and to libraries around the world through the IES. GPO currently distributes tangible journals through a system of trusted repositories that manage, maintain, and provide permanent public access to the journals. Under this system, the number of copies of the publication that are disseminated around the country and around the world strengthens permanent public access.

While GPO's mission includes permanent public access, many Federal agency publishers' missions do not. In order to make room for new issues and volumes, agency publishers often overwrite or remove old e-journal content. This practice frequently leads to the disappearance of Government e-journals from the Web. GPO and its library partners have identified a need for GPO to take physical custody of the "bits and bytes" that make up Government e-journals before they are removed by agency publishers.

### What is LOCKSS?

LOCKSS (for "Lots of Copies Keep Stuff Safe") is open source software that provides institutions with a way to collect, store, and preserve access to their own, local copy of content. LOCKSS was developed by Stanford University, and it is currently maintained by the Stanford University LOCKSS Program Management Office with support from the LOCKSS Alliance. LOCKSS runs on standard desktop hardware and requires minimal technical administration. Once installed, the LOCKSS software converts a personal computer into a digital preservation box that creates low-cost, persistent, accessible copies of e-journal content as it is published. The accuracy and completeness of content stored in a LOCKSS box is assured through a robust and secure, peer-to-peer polling and reputation system. A LOCKSS box performs the following four functions:



- \* It collects newly published content from the target e-journals using a Web crawler similar to those used by search engines.
- \* It continually compares the content it has collected with the same content collected by other boxes, and repairs any differences.
- \* It acts as a Web proxy or cache, providing browsers in the institution's community with access to the publisher's content or the preserved content as appropriate.
- \* It provides a Web-based administrative interface that allows the institution staff to target new journals for preservation, monitor the state of the journals being preserved, and control access to the preserved journals.

## Collecting

Before LOCKSS boxes can preserve a journal, two things have to happen:

- \* The publisher has to give permission for the LOCKSS system to collect and preserve the journal. They do this by adding a page to the journal's Web site containing a permission statement and links to the issues of the journal as they are published.
- \* The LOCKSS box has to know where to find this page, how far to follow the chains of Web links so that it doesn't crawl off the edge of the journal and try to collect the whole Web, some bibliographic information, and so on. In order to add new publishing platforms, the LOCKSS system provides a fill-in-the-blanks tool that a librarian or administrator can use to collect this information and test that it is correct. The information is then saved in a file (the LOCKSS plug-in) and added to the publisher's Web site or to some other plug-in repository, so that it is available to all LOCKSS systems.

## Preserving and Auditing

The LOCKSS boxes at libraries around the world use the Internet to audit, continually but very slowly, the content they are preserving. At intervals boxes take part in polls, voting on the digest of some part of the content they have in common. If the content in one box is damaged or incomplete that box will lose the poll, and it can repair the content from other boxes. This cooperation between the boxes avoids the need to back them up individually. It also provides unambiguous reassurance that the system is performing its function and that the correct content will be available to readers when they try to access it. The more organizations that preserve given content, the stronger the guarantee they each get of continued access.

## Providing Access

LOCKSS boxes provide transparent access to the content they preserve. Institutions often run Web proxies, to allow off-campus users to access their journal subscriptions, and Web caches, to reduce the bandwidth cost of providing Web access to their community. Their LOCKSS box integrates with these systems, intercepting requests from the community's browsers to the journals being preserved. When a request for a page from a preserved journal arrives, it is first forwarded to the publisher. If the publisher returns content, that is what the browser gets. Otherwise the browser gets the preserved copy.

## Administering



Staff administer their LOCKSS box via a Web user interface. It allows for targeting the appliance to preserve new journals, monitoring the preservation of existing journals, controlling access to the box and other functions.

Additional information about LOCKSS is available from the [Stanford University LOCKSS Web site](#).

## Pilot Project Description

GPO received numerous requests from research institutions, universities, depository libraries, and other Federal Government agencies to investigate using LOCKSS as a means to manage, disseminate, and preserve access to Web-based Federal Government e-journals that are within the scope of the FDLP and the IES. As a result, GPO conducted a 12 month pilot to make Federal Government e-journals available to select pilot libraries that are operating LOCKSS boxes. The following elements were used as "measures of success" for the pilot:

- \* GPO developed a list of 10 e-journals for inclusion in the pilot. Selection criteria included the following: disseminated via the FDLP, disseminated via the IES, selected by over 600 Federal depository libraries, and available in electronic format from a Federal agency Web site.
- \* GPO harvested 10 e-journals from Federal agency Web sites and added them to a secure folder on a GPO Web server.
- \* GPO developed plug-ins for all harvested pilot e-journals.
- \* Stanford tested and disseminated plug-ins for all harvested pilot e-journals.
- \* GPO developed publisher manifests and directory structures for all harvested e-journals.
  
- \* GPO provided access, via Pilot Partner LOCKSS boxes, to 10 e-journals stored on a GPO Web server.
- \* GPO maintained 10 e-journals stored on a GPO Web server. This included re-harvesting and updating content from agency Web sites in coordination with the release schedules for the pilot e-journals.
- \* Library Pilot Partners installed plug-ins and maintained a LOCKSS box for the duration of the pilot without exceeding estimated time and resource commitments.
- \* Library Pilot Partners used LOCKSS to harvest 10 e-journals from a secure GPO Web server without exceeding estimated time and resource commitments.
- \* GPO, Stanford, and Library Pilot Partners gathered data about project costs, operational experiences, technical experiences, quality control issues, and other matters related to electronic depository and IES distribution via LOCKSS.
- \* GPO, Stanford, and Library Pilot Partners developed, executed, and evaluated at least three real world scenarios.
- \* GPO evaluated the pilot.

## Pilot Partners

- \* Alaska State Library
- \* Arizona State University
- \* Brigham Young University
- \* Columbia University
- \* Dartmouth College
- \* Deutsche Bibliothek
- \* Georgetown University



- \* Georgia Tech
- \* Indiana University
- \* Library of Congress
- \* National Agricultural Library
- \* North Carolina State University
- \* Portland State University
- \* Rice University
- \* Stanford University
- \* University of Connecticut
- \* University of Kentucky Libraries
- \* University of Tennessee
- \* University of Utah
- \* University of Wisconsin-Madison
- \* U.S. Government Printing Office
- \* Yale University Law Library

## **LOCKSS Pilot e-Journals**

- \* Treasury Bulletin
- \* Social Security Bulletin
- \* Journal of Research of the National Institute of Standards and Technology
- \* Humanities
- \* Survey of Current Business
- \* Monthly Labor Review
- \* Monthly Energy Review
- \* FBI Law Enforcement Bulletin
- \* Amber Waves
- \* Environmental Health Perspectives

## **Pilot Project Report**

[Executive Summary](#)



[Report](#)

